Identification of Radicalization using Big data Analytics in Social Networks

SHORT SYNOPSIS

Submitted in partial fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

Manav Rachna International Institute of Research and Studies
(Deemed to be University)

by

Vandna Batra
15/Ph.D/017

Under the supervision of
Dr. Suresh Kumar
Professor, Faculty of Engineering & Technology

FACULTY OF ENGINEERING & TECHNOLOGY

MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH & STUDIES

SECTOR-43, SURAJKUND –DELHI ROAD, FARIDABAD – 121001

FACULTY OF ENGINEERING & TECHNOLOGY

Certificate

I hereby declare that the short synopsis "Title: Identification of Radicalization using Big data Analytics in Social Networks" being submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Computer Science & Engineering under faculty of Engineering & Technology of Manav Rachna International Institute of Research & Studies, Faridabad during the academic year 2019- 2020

It is a bona fide record of original work carried out under the supervision of Dr Suresh Kumar and has not been submitted in full or part to any other university or institute for degree or diploma. Further, it is certified that short synopsis is plagiarism free and proof has been submitted.

(Vandna Batra)
15/Ph.D/017

Signature of the Supervisor
Dr. Suresh Kumar
Professor, Faculty of Engineering & Technology

(Chairman, DRC)

Chairperson, CDP

ABSTRACT

Online radicalization or cyber hate extremism has become one of the threats to the society. Social media has been used as a major source for conducting radical activities by the extremists. Twitter has gain much more attention from the rest of the social media sites due to its popularity among users and ease of communication. Manual identification of radical tweets is nearly impossible due to presence of large amount of unstructured data. Automatic identification is also a difficult task as tweets not only contains text but also contains URLs, hashtags, retweets and so on. So it requires applying semi supervised learning techniques to identify the hidden patterns containing radical contents. In the existing work authors uses existing machine based classifiers along with manual intervention for detecting the radical groups. Also these radical groups are dynamic in nature. Due to the dynamic nature of such groups some members of the group may not be detected and radical activities are still going on. The research plan will be to filter out the hate, war, terrorist attacks related tweets from all the tweets posted on the Twitter. With the help of classifiers radical groups will be identified. Dynamic graph formation will also be done to reveal the interactions between different groups and shifting of members from one group to another with time. Many accounts related to terrorist have been banned by the government agencies but still radical activities are going on through unrevealed accounts which are to identify.

Keywords: Online radicalization, Twitter, radical, classifier, Dynamic graph

CONTENTS

S. No.	Description	Page No.
1	Introduction	1-4
	1.1 Sources of Big data	
	1.2 Social Network Analyses (SNA)	
	1.3 Applications of SNA	
2	Literature Review	5-8
3	Description of Broad Area	9-12
4	Objectives of the Study	13
5	Methodology to be adopted	14
6	Expected Outcome of the Research	15
7	References	16

1. Introduction

Data have been increasing at a very pace rate due to the abundance of sources. Volume of information and the decent variety of sources have encountered exponential development, making new specialized and application challenge. These led to the advent of new term called big data Yaqoob, I., Hashem, I.A.T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N.B. and Vasilakos, A.V. (2016). Big data is referred as a term for the gathering of informational collections that are so huge, complex with a wide diversity of types to such an extent that it ends up hard to store or process utilizing conventional data management tools.

1.1 Sources of Big data

1. Social media sites: The biggest source for big data is social media sites. They generate huge amount of unstructured data at an extremely rapid rate. The most popular social media sites are Facebook (through images, posts, likes, comments, videos etc.), Twitter (tweets, retweets, followers), LinkedIn, Instagram,Tumblr, Flickr and so on.
2. Data generated through transactions
3. Date generated through sensors that are used to gather climates, traffic, flights etc.
4. Internet of Things (IoT): IoT data are generated by smart TVs', mobile phones, almost every mobile phone is having GPS devices, smart watches, car window sensors and so on.

Social media has turned out to explode the big data era Bello-Orgaz, G., Jung, J.J. and Camacho, D., (2016). Social media data has been generated by various popular sites and internet applications. These are Facebook, Twitter, Tumblr, YouTube, LinkedIn, Instagram. Today, everyone lives in an associated world in which communication networks are interwoven with day by day life. These social networking sites have made another age of individuals sort of another general public who are showing intense interest and eager in interacting, sharing, and collaborating with the other people all around the world Bello-Orgaz, G., Jung, J.J. and Camacho, D., (2016).

Social Network provides web based services that allow individuals to create a profile within a province so that they can connect with other users within the network. Such platforms have been used by people for communication and also the data gathered from these sites have been

publicize by many diverse areas such as e-business, e-travel, education, health, political trending and forecasting. It not only changes the way we communicate but also help the business owners. Social networking influences business of all sizes in various perspectives. Due to the effect of social media every organization is now using them to publicize their products, brands and also using new approach to deal with their customers as it act as a low cost form of marketing. Thus with the help of social networking sites they are extending their business.

Social networks are an influential asset for spreading information, thoughts and similarly responsible for spreading false information, propaganda and so on. Therefore exploring these networks are of a great concern for the government law enforcement agencies, crime detection and cyber security departments.

Social media sites are carrying a lot of information; we need to store such data to gain insights out of it. Also the rate at which data is generated by social media sites is increasing rapidly and is mostly unstructured data. The data gathered has no use till we dig out some useful information out of them. So it requires the analysis of social media data.

Social media data analysis requires:

1. Intense analysis considering the actual data, perspective of the data, sentiment so as to provide deeper information.
2. Time consideration is also there as so as to extract the knowledge within the time before its importance goes off.
3. Network analysis to know how data is propagated, how an activity spreads over the network and how the network grows up.

1.2 Social Network Analysis

Social network analysis (SNA) is the procedure for identifying the connections, measurement of relationships between people or groups through the use of networks and graph. Social network is modeled by a graph in which nodes are the users and edges are the connection between different users. Using SNA one can visualize the data in order to reveal the patterns in the relationship.

1.3 Applications of Social Network Analysis

There are various applications of SNA, some of which have been discussed.

1. Sentiment Analysis

Sentiment Analysis refers to the overall mining of the data in order to determine whether the sentiment in the text is positive, negative or neutral. It provides a great knowledge to decision support system of an organization. With the increase in the use of social networking sites sentiment analysis has gained more value. Many tools and techniques have been provided by machine learning and Natural language processing that works with large volume of social media data to extracts the sentiments. With the help of social sentiment analysis can help the marketers to gain deeper context about their topic, brand or an advertisement.

2. Online Reputation Monitoring

Online reputation monitoring means finding others view point; positive or negative about a person, a brand or an organization over the internet community such as search engines, feedback forms, social networking sites etc. and try to solve the issues that may harm individual or brand reputation.

Every company is now doing social media monitoring in order to collect all the reviews about their products from blogs, tweets, likes and comments on Facebook and so on, and then processing them. Social Network Analysis can thus anticipate, and control the potential crisis of an individual or a brand reputation. It can be done by eliminating all the negative information about your band over the social media and working upon the feedbacks given by the users in order to improve the brand name and also strengthening the trust of customers upon you.

3. Online Fraud Detection

Online fraud is a fraud that is carried out using the internet. It can be in terms of finance as well as identity steal. Deceitful activities have been committed through emails, messages, web sites etc where fake information has been provided to the user in order to get the bank account details to steal out the money or sometimes stealing the identity of user to conduct any illegal activity. Online fraud keeps on increasing day by day; it affects millions of people and also led to tremendous financial loss to businesses.

Social media sites have been used by almost every user in the world. Fraudsters can reach out a large number of audiences through them and can make their messages look genuine that it becomes difficult to detect the fraud activities. Social network analysis has been used by various

business sectors to combat online fraud. SNA can be applied to develop graph structure through which suspicious individuals, their relationships with other person that may change over time can be detected.

4. Fake Review Detection

Social media sites act as a main and low cost form of marketing and advertisements done there influences people more. Mostly people buy the product after seeing its reviews. People rely a lot on the written reviews that may influence their decision about the purchase of the product. Reviews can be positive or negative. Good reviews definitely increase the sales of a product. But there are also some reviews that are not genuine called fake reviews. These reviews are purposely made to present a false impression to customers at the time of buying the product. Such reviews are difficult to detect as they are done by individuals who create a fake account to post the fake reviews in replace of money. Analysis of network is required to detect such accounts.

5. Online Radicalization

Social media offers an easy environment and wide network for exchanging information. Due to its low publication barrier it has also been used by the activist groups for conducting illegal activities. Extremists use this platform to spread out their propaganda called online radicalization. Terrorist organization called radical groups uses both websites as well as social media platforms such as Twitter, Facebook, Youtube, Tumblr etc. to exchange their ideas with likeminded people, plan out attacks, share videos on how to create bombs, also provide religious explanations in order to involve new people into their groups. These groups need to be identified in order to stop such activities.

2. Literature Review

Data is growing at a very fast rate. The main sources contributing the increase of data are IoT, social media sites, sensors etc. Such huge amount of data called big data generates various new challenges.

Chen, C.P. and Zhang, C.Y., (2014) discussed many challenges related to acquiring data, storing data, analyzing and visualizing the data. The author proposed several new tools and techniques to deal up with those challenges.

Yaqoob, I., Hashem, I.A.T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N.B. and Vasilakos, A.V., (2016) describes data processing tools and technologies based on both batch based and stream based data. Out of them stream data processing is more prominent. Main source of generating streaming data is social media. Within few seconds billions of data has been generated that are dynamic in nature. Thus to handle the large amount of streamed data existing data mining algorithms, visualization methods, machine learning algorithms all needed to be scaled up to handle such large amount of data which is a challenging task.

Bello-Orgaz, G., Jung, J.J. and Camacho, D., (2016) discussed that social media sites is the main source of generating big data and it creates a new society of people who are passionate to share their pictures, videos, exchanging ideas with peoples throughout the world. The data generated have been used by health, forecasting, e–business, tourism aim to extend their market.
Social media created new challenges such as:

- The data generated are in unstructured form so difficult to analyze such data sets.
- Streaming data are dynamically changing thus it also challenging to analyze them.
- Privacy and security issue: Sharing geographical location, personal information etc. on social media may be used by some illegal persons to harm you.

Social networks have been used by businesses to manage their brand reputation, get the reviews about their products. Social networks due to its ease of accessibility and reaching out a mass audience, has been used for spreading information, thoughts and similarly responsible for spreading false information, propaganda and so on. They have been used by different society

people, whose motive is to enhance cybercrimes, conducting terrorist activities. Therefore analysis of social network has been required to detect such activities.

Correa, D. and Sureka, A., (2013) mentioned that Cyber-Terrorism or Extremism is boundless and has turned into a noteworthy and developing worry to the general public, governments and law requirement offices around the globe. Therefore automatic detection of such tasks on the Internet is the key requirement. Author gives a survey of automated solutions to detect cyber terrorism and their limitations that provide a future scope for the improvements.

Wadhwa, P. and Bhatia, M.P.S., (2013) discusses that over the internet social media sites are the main source for conducting illegal activities. Violent extremist groups are diligently doing various criminal tasks through these sites. Social media sites such as Twitter, Facebook, Instagram, Tumblr, LinkedIn, YouTube etc. all have been used by such groups to promote their ideas, involving new people in their group and planning terrorist activities. Contents in the messages shared are in multiple languages and there may be link between different groups. Discovering such things is a challenging task. Author works on Twitter data set for one day and search for the tweets with the predefined hashtag by using semi-supervised n-gram approach taking n<3. The limitations in this paper are that the author considers only text and ignore the hyperlinks but is has been seen that text may not contain any information related to extremism but the links in the text may open sites having misleading content. Also the author filters the dataset for English language only while the contents are multilingual.

Agarwal, S. and Sureka, A., **(2015)** describes that social media platforms specially twitter have been used for conducting terrorist and other criminal activities by the extremist groups as tweets are short messages that can be posted without any barrier and has a wide accessibility among audience. The main two research areas have been discussed, that are detecting and analyzing such content automatically. The author configures that to analyze multilingual text on twitter dynamic clustering; Logistic Regression and Graph Traversal techniques are the mainly preferred techniques for detecting different hidden groups and their linking with other groups.

Chatfield, A.T., Reddick, C.G. and Brajawidagda, U., (2015) analyzes social network and discovers that Twitter has been widely used by Islamic terrorist networks to spread their propaganda and radicalization**.** Author works on one of the renowned account ISIS disseminator on Twitter to detect such activities. He also discovers that network of terrorist have been continuously changing due to changes in policies and for enrollment purposes as well. So it requires more research to detect all such terrorist networks on the twitter.

Agarwal, S. and Sureka, A., (2015) works on twitter data set. Author extract tweets on some predefined hashtags related to terrorism and generates some more hashtags out of them. Then extract tweets with them also thus extended the list of hashtags and training data set. After that author applies preprocessing techniques to remove urls, retweets, hashtags. SVM and KNN classifiers are then applied to filter out hate promoting tweets and accuracy has been calculated using F-score with a value of 0.83 and 0.6 respectively. Author concludes that the presence of war related, religious, nasty words, negative emotions are the main signs of indicating hatred upholding tweets.

.

Wadhwa, P. and Bhatia, M.P.S., (2015) made a data set comprising of tweets related to hashtags such as ‘‘al-Qaeda’’, ‘‘jihadist’’, ‘‘radicalism’’ that were extracted from Twitter for a period of three days. The messages caught were preprocessed for the expulsion of end words, punctuations; extra spaces and other dissipate words. After that dynamic network graphs were created. With the help of such interactions graphs most conspicuous users were detected. But it has been seen that the most conspicuous users that are discovered out through interaction graphs may not be the important users and they may be lying at the boundary of the graphs. So it requires more work to be done to investigate such graphs deeply.

Ashcroft, M., Fisher, A., Kaati, L., Omer, E. and Prucha, N., (2015) focused on the circulation of messages related to jihad on Twitter. They have gathered a lot of tweets having hashtags identified with jihad and also with their corresponding word in Arabic. They applied three classes of features stylometric features for searching on the basis of more frequent words, time based features to search for the day of a week and time at which groups are more active and lastly sentiment features to determine the nature of text towards certain topics. Three classification algorithms have been used by them. These are SVM, Naïve Bayes and Adaboost.

Author applies semi automatic techniques for feature extraction and specifies Friday as a critical day to spread radical tweets.

Lara-Cabrera, R., Gonzalez-Pardo, A. and Camacho, D., (2017) works over 17000 tweets taken form a dataset in Kaggle for the pro-ISIS account on Twitter and measures the risk of individual to get radicalized. For this purpose author uses a 5 set of indicators. These are: Individuals' attitude (frustrated, introvert), impression of separation of being Muslim, communicating negative thoughts for Western Society and positive towards jihadist. Pairwise correlation has been done between different indicators and strong correlations have been detected between majorities of indicators. Once such individuals have been identified, some actions can be taken to avoid them to become radicalize or jihadist.

Fernandez, M., Asif, M. and Alani, H., (2018) did analysis to determine the root cause of radicalization which can be at micro, meso and macro level. Author works on two publically datasets available on Kaggle and apply multiple classifiers on the dataset (J48, Naïve Bayes and Log R) to anticipate radicalization impact using a set of known radicalized keywords.

Sapountzi, A. and Psannis, K.E., (2018) provides a survey of various social network analysis tools that uses graph algorithms for analysis and visualization tasks and they doesn't require any coding. The tools that have been discussed are:

- NodeXL
- Gephi
- NetworKit
- Pajek
- Statnet

Comparisons of these tools have been done on various parameters such as platform used, analysis needs of users and analytical capabilities of tools.

3. Description of Broad Area

Social media come into sight as one of the most popular way of communication. In today's society people lives in a connected world where they communicates with other people throughout the world through this social network. It has become a daily part of their lives. The most popular websites for communication are Facebook, Twitter, LinkedIn, YouTube, Instagram, Tumblr etc. Large amount of data has been generated by these sites and it becomes a challenge to store and analyze such data Bello-Orgaz, G., Jung, J.J. and Camacho, D., (2016).

Table I. Social media websites and mobile application statistics

Social media sites	**Statistics**
Twitter(Jan 2018)	Number of Twitter Daily Active Users: 100 million Number of Tweets sent per Day: 500 million Total Number of Monthly Active Twitter Users: 330 million
Facebook(Jan 2018)	No. of daily active users: 1.40 billion Photo uploads per day: 300 million No. of comments per minute: 510,000
LinkedIn(Jan 2018)	Total Number of LinkedIn Users: 500 million Total Number of Monthly Active LinkedIn Users: 250 million Number of LinkedIn new Members per Second:2
Tumblr(Feb 2018)	Total blog accounts : 396.8 million Monthly users: 550 million Total posts: 158.1 billion No. of daily posts: 32.3 million
YouTube(Jan 2018)	Number of videos uploaded in a minute:300 hours Number of videos watched in a single day: approx. 5 billion Total number of hours of video watched each month: 3.25 billion.
Instagram(Dec 2017)	Total Number of Monthly Active Instagram Users: 800 million Total Number of Daily Active Instagram Users: 500 million Number of Instagram Likes per day: 4.2 billion

These sites has made another age of individuals sort of another general public who are indicating exceptional intrigue and energetic in about connecting, sharing, and teaming up with the other

individuals all around the globe. The data accumulated from here has been plug by a wide range of regions, for example, e-business, e-travel industry, training, healthcare, political inclining and anticipating Yaqoob, I., Hashem, I.A.T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N.B. and Vasilakos, A.V. (2016).

Social media sites have gained as a source to ease the communication by providing simple, quick and inexpensive way for exchanging ideas. Due to this ease provided it has been used as a major media by other society people having misleading nature. They promote criminal activities, terror attacks, attracting new audience around the world through this social media sites. Social media is being used by such people for spreading their activities, involving more and more people thus extending their group. These extremist groups lure new people to enter in their group either by money or by disrupting their minds in terms of religion. Agarwal, S. and Sureka, A., (2015) gives as overview how extremist people use social media for conducting terrorist activities. They uses such platform because it provides an easy way of communication and posting messages that has not been monitored by any government agency and has no regulations.

Social Media Pros and Cons

- **Pros:**
  - Effective way of communication
  - Let the peoples connect worldwide
  - One can reach out a wide range of audience to share something.
  - Information collected can be used for e-business, e-tourism, and e-education.
  - Used by business owners for expansion of their business as it act as a low cost form of marketing.
- **Cons:**
  - Privacy issues
    - Privacy of users need to be preserved as the risk of leakage of private information of user through social media sites is growing and also preventing it from any illegal usage.
  - It can be used to spread rumors at mass level.

- Used for illegal activities such as: terrorist attacks planning, provoking people against racisms, nationalism, and on religious terms Agarwal, S. and Sureka, A., (2015).
- Terrorist organization uses social media for recruiting new members by targeting the audience at individual level, group level and global level and thus spreading their propaganda Wadhwa, P. and Bhatia, M.P.S., (2013).
- This led to invent of a new term called Online Radicalization Wadhwa, P. and Bhatia, M.P.S., (2013). Online radicalization also called cyber hate extremism consists of using Internet as a medium by extremists and terrorists for conducting malicious activities. Cyber fanaticism, cyber hate propaganda (Online radicalization) has emerged as one of the high-flying risk to the society, governments and law requirement organizations.

Various social networking sites have been used for communication out of them Twitter is much more popular as it offers users to post short messages called tweets. Twitter creates more than 500 million tweets each day.

It has been widely used as a platform for online radicalizations promoting extremist content due to its popularity among users. It is nearly impossible to manually identify tweets done by such users and radical groups Agarwal, S. and Sureka, A., (2015). Tweets comprise of short content (limit of 140 characters) and also contain text having spelling mistakes, wrong grammar, short forms, and hashtags. Therefore automatic identification of such tweets is also a complicated task.

There are also various challenges associated:

1. Radical or extremist groups are dynamic in nature i.e. members of the group continue to shift from one group to another with time and thus requiring time based collection of information and analysis.
2. Messages shared by such members have hidden information, therefore sentiment analysis need to be applied to reveal hidden patterns.
3. Not only text messages one has to deal with unstructured data as well. Illegal information may be embedded in images, video. Sometimes even text messages do not reveal any criminal act due to some hidden meaning or some coded words. Also some hyperlinks are there in the text that may deviate towards extremist activities.
4. Contents mentioned in the text are multilingual. Analyzing such data is also challenging.

5. Wadhwa, P. and Bhatia, M.P.S., (2013) explains that there might be link between different radical groups. Some people may be active in one group but in some situation or time or day they might shift to another group or cluster showing their relationship with that group also.

It is nearly impossible to manually identify tweets done by such users and radical groups. Tweets comprise of short content (limit of 140 characters) and also contain text having spelling mistakes, wrong grammar, short forms, and hashtags. Therefore automatic identification of such tweets is also a complicated task. There is a lot of research going on to identify radical groups. There are many accounts related to terrorist groups that have been banned by Twitter but terrorist activities are still going on through unrevealed accounts. Keeping track of the contents shared by them is necessary to identify their planning in advance so that necessary actions can be taken place. There are a few undiscovered open doors in identifying misusing of open-source internet based life information which need to be work upon.

4. Objectives of the Study

1. To study datasets related to radicalization.
2. To apply existing machine based classifiers for detecting hate and extremism contents in the dataset and to compare accuracy results.
3. To identify the static and dynamic radical groups using clustering techniques and graph theory.
4. To evaluate new datasets and applying different techniques for the identification of the radical group activities.
5. To design a classifier for separating active groups by identification of the day and time frames.

5. Methodology to be adopted

Data will be fetched from Twitter. For analysis R tool will be used and also for development of new solution. After analysis graphic display will be shown again using R tool. Reason for choosing R tool as its open source software and provides a large collection of packages for statistical computing and graphics.

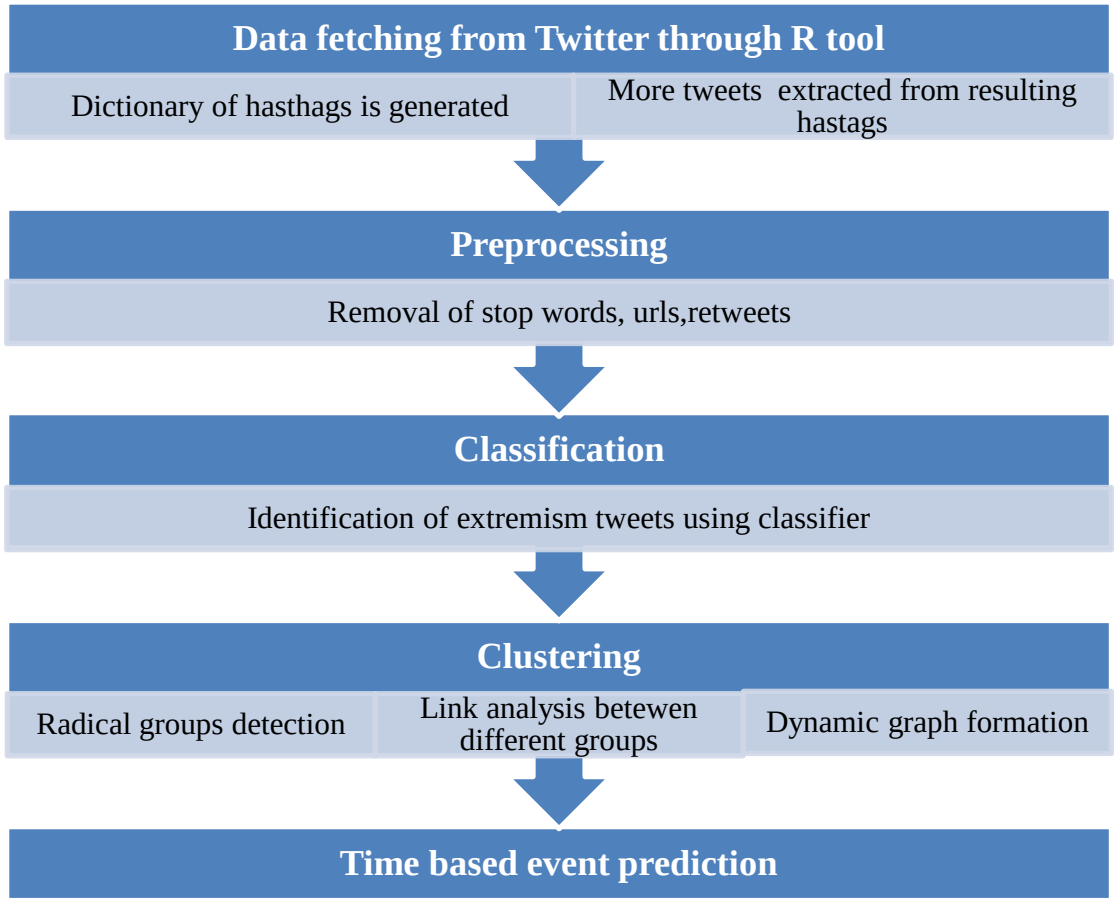

Figure2 Flow diagram of research plan

6. Expected Outcome of the Research

1. Dictionary of known hashtags containing hate and extremist content will be created. Tweets related to such seeds hashtag will be fetched through Twitter using R tool. List of hashtags will be extended from the data fetched and will be added to the dictionary. New tweets will be fetched.
2. Machine based classifiers will be applied to classify radical contents from the data fetched and accuracy of the classifiers will be calculated.
3. Comparative analysis will be done of the result generated.
4. Dynamic graph formations will be done to identify radical groups.
5. Time based predictions of radical activities and new group formation will be done.

References

Agarwal, S. and Sureka, A., 2015. Applying social media intelligence for predicting and identifying on-line radicalization and civil unrest oriented threats. *arXiv preprint arXiv:1511.06858*.

Agarwal, S. and Sureka, A., 2015,. Using knn and svm based one-class classifier for detecting online radicalization on twitter. In *International Conference on Distributed Computing and Internet Technology* pp. 431-442.

Ashcroft, M., Fisher, A., Kaati, L., Omer, E. and Prucha, N., 2015, September. Detecting jihadist messages on twitter. In *2015 European Intelligence and Security Informatics Conference* pp. 161-164.

Bello-Orgaz, G., Jung, J.J. and Camacho, D., 2016. Social big data: Recent achievements and new challenges. *Information Fusion*, *28*, pp.45-59.

Chatfield, A.T., Reddick, C.G. and Brajawidagda, U., 2015,. Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks. In *Proceedings of the 16th Annual International Conference on Digital Government Research* pp. 239-249.

Chen, C.P. and Zhang, C.Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, *275*, pp.314-347.

Fernandez, M., Asif, M. and Alani, H., 2018, May. Understanding the roots of radicalisation on twitter. In *Proceedings of the 10th ACM Conference on Web Science* pp. 1-10.

Han, J., Pei, J. and Kamber, M., 2011. *Data mining: concepts and techniques*.

Lara-Cabrera, R., Gonzalez-Pardo, A. and Camacho, D., 2017. Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in Twitter. *Future Generation Computer Systems*.

Sapountzi, A. and Psannis, K.E., 2018. Social networking data analysis tools & challenges. *Future Generation Computer Systems*, *86*, pp.893-913.

Sureka, A. and Agarwal, S., 2014,. Learning to classify hate and extremism promoting tweets. In *2014 IEEE Joint Intelligence and Security Informatics Conference* pp. 320-320.

Wadhwa, P. and Bhatia, M.P.S., 2013,. Tracking on-line radicalization using investigative data mining. In *2013 National Conference on Communications (NCC)* pp. 1-5.

Wadhwa, P. and Bhatia, M.P.S., 2015. An approach for dynamic identification of online radicalization in social networks. *Cybernetics and Systems*, *46*(8), pp.641-665.

Wei, Y., Singh, L. and Martin, S., 2016, August. Identification of extremism on twitter. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* pp. 1251-1255.

Agarwal, S. and Sureka, A., 2015. Applying social media intelligence for predicting and identifying on-line radicalization and civil unrest oriented threats. *arXiv preprint arXiv:1511.06858*.

MANAV RACHNA
vidyaprateekshan

Dr. O.P. Bhalla Central Library
Manav Rachna International Institute of Research and Studies
(A deemed-to-be-university under section 3 of UGC Act 1956)

SIMILARITY VERIFICATION CERTIFICATE

Title: Identification of Radicalization using Big Data Analytics in Social Networks

Researcher : Vandna Batra Registration Number: 15/Ph.D/a17

Supervisor: Dr. Suresh Kumar Department : Computer Science

This is to certify that the above mentioned Short Synopsis/~~Thesis~~ was scanned for similarity index through Turnitin Software on 17/9/19 withsubmission ID 11.7429.1631 and the similarity index was found to be 6 %.The complete report running into 20 Pages, 4372 Word count with File Size 338a KBis being forwarded to the Research Supervisor/DRC/DRB for review.

Dr. Vikram Kr. Sharma
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

Date: 20/9/19

University Librarian

The complete report of the above Short Synopsis/~~Thesis~~ has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/~~Thesis~~ may be considered for further processing

Vandna

Research Scholar

Research Supervisor

20-09-2019

Vandana Batra Short Synopsis Sep 17 2019

ORIGINALITY REPORT

Excluding matches less than 14 words

6%	4%	3%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	scitepress.org Internet Source	1%
2	oeconomica.uab.ro Internet Source	1%
3	Wadhwa, Pooja, and M.P.S Bhatia. "Discovering Hidden Networks in On-line Social Networks", International Journal of Intelligent Systems and Applications, 2014. Publication	1%
4	Submitted to University of Sheffield Student Paper	1%
5	blog.socialmediamanagement.net Internet Source	<1%
6	Submitted to University of Wales Swansea Student Paper	<1%
7	Kinza Zahra, Farooque Azam, Wasi Haider Butt, Fauqia Ilyas. "A Framework for User Characterization based on Tweets Using Machine Learning Algorithms", Proceedings of	<1%

Dr. Vikram Kr. Sharma
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

Vandana

20/09/2019

	the 2018 VII International Conference on Network, Communication and Computing - ICNCC 2018, 2018 Publication	
8	Swati Agarwal, Ashish Sureka, Vikram Goyal. "Chapter 2 Open Source Social Media Analytics for Intelligence and Security Informatics Applications", Springer Science and Business Media LLC, 2015 Publication	<1%
9	Submitted to SVKM International School Student Paper	<1%
10	Submitted to Bowling Green State University Student Paper	<1%

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		